REDUCED MEAN SQUARE ERROR ESTIMATION FOR SEVERAL PARAMETERS[1]

by

Michael D. Perlman

Technical Report No. 175

April 1972

University of Minnesota
Minneapolis, Minnesota

[1]This research was supported in part by U.S. Army Research Office grant number DA-ARO-D-31-124-71-G171.

Summary.

If t is an unbiased estimator of a vector parameter, a necessary and sufficient condition is given for the existence of a constant α, $0 < \alpha < 1$, such that αt has uniformly smaller mean square error than t in a strong sense. Applications are made to the estimation of several scale parameters and of a covariance matrix.

1. A Simple Improvement of an Unbiased Estimator.

Suppose t is an unbiased estimator of a real-valued parameter $\tau = \tau(\theta)$. Blight (1971) compared t with estimators of the form αt, where $0 < \alpha < 1$ is a constant. He showed that a sufficient condition (also necessary) for the existence of an estimator αt having uniformly smaller mean square error (m.s.e.) than t is that the quantity $\tau^2(\theta)/\mathrm{Var}_\theta(t)$ be bounded over the range of θ.

In this note we consider the multiparametric case. Suppose now that $t = (t_1, \ldots, t_p)$ is an unbiased estimator of the vector of unknown parameters $\tau = \tau(\theta) = (\tau_1(\theta), \ldots, \tau_p(\theta))$. Assuming that the covariance matrix $V = V(\theta) = \mathrm{cov}_\theta(t)$ is nonsingular, we show that <u>the boundedness of the scalar quantity $\tau(\theta)V^{-1}(\theta)\tau(\theta)'$ over the range of θ is necessary and sufficient for the existence of a constant α, $0 < \alpha < 1$, such that the vector estimator αt dominates t in the following strong sense</u>:

$$E_\theta(\alpha t - \tau)'(\alpha t - \tau) \leq E_\theta(t - \tau)'(t - \tau) \tag{1}$$

<u>for all</u> θ. We write $A \leq B$ for two $p \times p$ symmetric matrices to mean that $B - A$ is positive semidefinite. Condition (1) implies that for every linear combination $\sum_{i=1}^{p} c_i \tau_i$ of the unknown parameters, the estimator $\alpha \sum c_i t_i$ has m.s.e. no greater than that of the unbiased estimator $\sum c_i t_i$.

To prove the result, rewrite (1) as $\alpha^2 V + (1 - \alpha)^2 \tau'\tau \leq V$, or equivalently, $\tau'\tau \leq m\,V$ where $m = (1 + \alpha)/(1 - \alpha)$. If $V^{\frac{1}{2}}$ denotes any square root of the positive definite matrix V, then this last inequality is equivalent to

$$(\tau\, V^{-\frac{1}{2}})'(\tau\, V^{-\frac{1}{2}}) \leq m\, I \tag{2}$$

where I is the $p \times p$ identity matrix. By the Cauchy-Schwartz inequality, however, (2) is equivalent to the single scalar inequality $\tau V^{-1}\tau' \leq m$. Thus, if there exists $0 < \alpha < 1$ such that (1) holds for all θ, then $\tau V^{-1}\tau'$ is bounded. Conversely, if $\sup_\theta \tau V^{-1}\tau' \equiv m_0 < \infty$, then (1)is satisfied for any $0 < \alpha < 1$ such that $(1 + \alpha)/(1 - \alpha) \geq m_0$, or equivalently,

$$\alpha \geq (m_0 - 1)/(m_0 + 1). \tag{3}$$

Notice that equality cannot hold in (1); otherwise,it would also hold in (2), which is impossible since the left side of (2) is a matrix of rank one. Thus for α satisfying (3), αt is a <u>strictly</u> better estimator than t, and $\alpha \sum c_i t_i$ will have <u>strictly</u> smaller m.s.e. than $\sum c_i t_i$ for most linear combinations $\sum c_i \tau_i$.

Next, suppose that the underlying probability distribution is determined by a density or discrete mass function $f = f(x_1, \ldots, x_n; \theta)$. If $\theta = (\theta_1, \ldots, \theta_r)$ and if appropriate regularity conditions are satisfied, then the multiparametric version of the Cramer-Rao inequality (Rao (1965), p.265) states that $V(\theta) \geq T(\theta)J(\theta)^{-1}T(\theta)'$. Here $J(\theta)$ is the $r \times r$ information matrix with entries $E(-\partial^2 \log f/\partial\theta_i \partial\theta_j)$ and $T(\theta)$ is the $p \times r$ matrix of partial derivatives $\partial\tau_i/\partial\theta_j$. If $TJ^{-1}T'$ is nonsingular, then the C - R inequality implies that $\tau V^{-1}\tau \leq \tau(TJ^{-1}T')^{-1}\tau'$, so $m_0 \leq m_1 \equiv \sup_\theta \tau(TJ^{-1}T')^{-1}\tau'$. The quantity m_1 depends only on the family of distributions $f(x_1, \ldots, x_n; \theta)$ and the parameters $\tau_i(\theta)$ to be estimated, not on a particular statistic t. <u>Therefore</u>, $m_1 < \infty$ <u>implies that no unbiased estimator</u> t <u>can be admissible for estimating</u> τ: whenever $\alpha \geq (m_1 - 1)/(m_1 + 1)$, αt dominates t in the sense of (1).

2. Choice of the Reduction Factor α.

When $p = 1$, τ and V are scalars so for each fixed θ,

$$\alpha^* = \alpha^*(\theta) \equiv \tau^2/(\tau^2 + V) \tag{4}$$

minimizes $\alpha^2 V + (1 - \alpha)^2 \tau^2$, the left side of (1). If τ^2/V does not depend on θ, then α^* is the optimal value (Goodman(1953), Theorem 1). If $p > 1$, however, there does not exist a value α^* which minimizes the left side of (1) with respect to the partial ordering $\leq$, even for θ fixed. To see this, use the fact that there exists a nonsingular $p \times p$ matrix $F = F(\theta)$ such that $V = F'F$ and $\tau'\tau = F'DF$, where D is the $p \times p$ diagonal matrix $\mathrm{diag}\{\tau V^{-1}\tau', 0, \ldots, 0\}$(Anderson(1958), Theorem 3, p.341). Thus

$$\alpha^2 V + (1-\alpha)^2 \tau'\tau = F'\mathrm{diag}\{\alpha^2 + (1-\alpha)^2 \tau V^{-1}\tau', \alpha^2, \ldots, \alpha^2\}F, \tag{5}$$

so the optimal value α^*, were it to exist, would have to simultaneously minimize $\alpha^2 + (1-\alpha)^2 \tau V^{-1}\tau'$ and α^2 over $0 < \alpha < 1$, which is impossible since $\tau V^{-1}\tau' > 0$.

The representation (5) does show that α should be chosen to satisfy $\alpha \leq \sup_\theta(\tau V^{-1}\tau')/(\tau V^{-1}\tau'+1) = m_0/m_0+1$ (compare this with (3) and (4)). Furthermore, (5) suggests that if one wishes to estimate only some, not all, linear combinations $\sum c_i \tau_i$, then, possibly, (3) should not be heeded, especially for large p. This is illustrated in the next section.

3. Applications.

The first two examples concern estimation of several scale parameters. Let $U = (U_1, \ldots, U_p)$ have a known distribution with $u_i \equiv EU_i \neq 0$ and $W \equiv \mathrm{cov}(U)$ nonsingular, and let $(X_1, \ldots, X_p) = (\theta_1 U_1, \ldots, \theta_p U_p)$, where

$\theta_1, \ldots, \theta_p$ are unknown scale parameters, $\theta_i > 0$. Then $t \equiv (X_1/u_1, \ldots, X_p/u_p)$ is an unbiased estimator of $\tau(\theta) \equiv (\theta_1, \ldots, \theta_p)$. Here $V(\theta) = D(\theta)WD(\theta)$, where $D(\theta) = \text{diag}\{\theta_1/u_1, \ldots, \theta_p/u_p\}$. Therefore, setting $u = (u_1, \ldots, u_p)$, one sees that $\tau V^{-1}\tau' = uW^{-1}u'$ does not depend on θ, hence is bounded. Thus $m_0 = uW^{-1}u'$, and for any α satisfying (3), the estimator αt dominates t in the sense of (1).

Example 1.

Let $X_1, \ldots, X_p$ be independent random variables with X_i uniformly distributed on the interval $(0, \theta_i)$. Then $t = (2X_1, \ldots, 2X_p)$ is an unbiased estimator of $(\theta_1, \ldots, \theta_p)$. Here each U_i is uniformly distributed on $(0, 1)$, so $u = (\frac{1}{2}, \ldots, \frac{1}{2})$, $W=(1/12)I$, and $m_0 = uW^{-1}u' = 3p$. Thus αt dominates t whenever $\alpha \geq (3p-1)/(3p+1)$.

Example 2.

Let $\{Z_{ij} \mid j = 1, \ldots, N_i \, ; \, i = 1, \ldots, p\}$ be independent observations from p normal populations, with $Z_{ij} \sim N(\mu_i, \theta_i)$, and let $X_i = \sum_{j=1}^{N_i}(Z_{ij} - \bar{Z}_i)^2$. Then $X_i \sim \theta_i \chi^2_{n_i}$ where $n_i = N_i - 1$, and $t = (X_1/n_1, \ldots, X_p/n_p)$ is an unbiased estimator of $(\theta_1, \ldots, \theta_p)$. Here $U = (\chi^2_{n_1}, \ldots, \chi^2_{n_p})$ is a vector of chi-square variates, so $u = (n_1, \ldots, n_p)$ and $W = \text{diag}\{2n_1, \ldots, 2n_p\}$. Hence $m_0 = uW^{-1}u' = k/2$. where $k = \sum n_i$ is the total degrees of freedom. Therefore αt dominates t whenever $\alpha \geq (k-2)/(k+2)$.

The next example concerns estimation of a covariance matrix.

Example 3.

Let $Y_1, \ldots, Y_N$ (each a $1 \times p$ row vector) be independent observations from a p-variate normal distribution with mean vector μ and nonsingular covariance matrix Σ. Then $S \equiv \sum_{i=1}^{N}(Y_i - \bar{Y})'(Y_i - \bar{Y})$ has a Wishart distribution $W(\Sigma, n)$ where $n = N-1$ (Anderson(1958), Chapter 7). The matrix

S/n is an unbiased estimator of Σ. Since Σ is symmetric we are estimating $q \equiv p(p+1)/2$ parameters. Expressed in vector form $t \equiv n^{-1}(s_{11},\ldots, s_{pp}, s_{12},\ldots, s_{1p},\ldots, s_{p-1,p})$ (a vector of length q) is an unbiased estimator of $\tau \equiv (\sigma_{11},\ldots, \sigma_{pp}, \sigma_{12},\ldots, \sigma_{1p},\ldots, \sigma_{p-1,p})$, where $S = (s_{ij})$, $\Sigma = (\sigma_{ij})$. To evaluate $\tau V^{-1}\tau'$, write $S = \Sigma^{\frac{1}{2}} S^* \Sigma^{\frac{1}{2}}$, where $S^* \equiv \Sigma^{-\frac{1}{2}} S \Sigma^{-\frac{1}{2}} \sim W(I, n)$. Thus S is a linear function of $S^* = (s^*_{ij})$, so t must be a linear function of $t^* \equiv n^{-1}(s^*_{11},\ldots, s^*_{pp},\ldots, s^*_{p-1,p})$, say $t = t^*A$, where the $q \times q$ nonsingular matrix A depends on Σ. Therefore $\tau = Et = (Et^*)A$ and $V = \text{cov}(t) = A'\text{cov}(t^*)A$, so $\tau V^{-1}\tau' = (Et^*)\{\text{cov}(t^*)\}^{-1}(Et^*)'$ does not depend on Σ. Now, $Et^* = (1,\ldots, 1,0,\ldots, 0)$ (p ones followed by $p(p-1)/2$ zeroes) and, applying equations (14) and (15) on p. 161 of Anderson (1958), $\text{cov}(t^*) = n^{-1}\text{diag}\{2,\ldots,2,1,\ldots,1\}$($p$ twos followed by $p(p-1)/2$ ones). Hence $\tau V^{-1}\tau' = np/2 = m_0$, so $\alpha(S/n)$ is a uniformly better estimator of Σ than S/n in the sense of (1) whenever $\alpha \geq (np-2)/(np+2)$.

In all three examples $m_0 \to \infty$ as $p \to \infty$, so the requirement (3), while guaranteeing that αt is a uniform improvement over t in the sense of (1), gives only a very small improvement for large p. If one wants to estimate only certain linear combinations $\sum c_i\tau_i$, then for moderate or large p it may be preferable to choose $\alpha < (m_0-1)/(m_0+1)$. In Example 1, to estimate each θ_i individually $E\{\alpha(2X_i) - \theta_i\}^2$ is minimized when $\alpha = 3/4$, which is less than $(3p-1)/(3p+1)$ whenever $p > 2$. In Example 2, assuming for simplicity that $n_1 = \ldots = n_p = n$, $E\{\alpha(X_i/n) - \theta_i\}^2$ is minimized when $\alpha = n/n+2$, which is less than $(k-2)/(k+2) = (np-2)/(np+2)$ when $p > 4$. In Example 3, $E\{\alpha(s_{ii}/n) - \theta_i\}^2$ is also minimized when $\alpha = n/n+2$; while for $i < j$, $E\{\alpha(s_{ij}/n) - \sigma_{ij}\}^2$

is minimized when $\alpha = n\rho_{ij}^2/\{(n+1)\rho_{ij}^2 + 1\}$ (where $\rho_{ij}^2 = \sigma_{ij}^2/\sigma_{ii}\sigma_{jj}$) which depends on the unknown parameters but which is never greater than $n/n+2$. Here again, $n/n+2 < (np-2)/(np+2)$ when $p > 4$. Thus in all three examples, not heeding (3) enables one to further reduce the m.s.e. for those linear combinations $\sum c_i\tau_i$ of particular interest. If, however, one wishes to estimate not only each τ_i individually but also contrasts of the form $\tau_i - \tau_j$ and possibly other linear combinations, then condition (3) should be maintained.

REFERENCES

Anderson, T. W. (1958). An Introduction to Multivariate Statistical Analysis. New York: Wiley.

Blight, B. J. N. (1971). Some general results on reduced mean square error estimation. The American Statistician **25**, No. 3, 24-25.

Goodman, L. (1953). A simple method for improving some estimators. Ann. Math. Statist. **24**, 114-117.

Rao, C. R. (1965). Linear Statistical Inference and its Applications. New York: Wiley.